Final Technical Report

Project Title: Information Theoretic Causal Coordination

AFOSR Award Number: AF FA9550-10-1-0345
Reporting Period: July 15, 2010 - July 14, 2013

Principle Investigator: Negar Kiyavash, University of Illinois at Urbana-Champaign

Submitted to: Dr. Tristan Nguyen, Air Force Office of Scientific Research (AFOSR/RSL)

1 Executive Summary

Automated decision-making over large-scale distributed systems in the presence of uncertainty and incomplete information (or purposefully inaccurate information) is a formidable task beyond the research capability of any single field. Specifically, any successful approach must view information from different perspectives, including (i) probability distributions; (ii) discrete numeric values; (iii) linguistic statements, and combinations thereof. The information sources can be physical sensors or humans such as experts in a given area. Therefore for fusion and inference on such diverse types of information a multitude of different methods must be explored. These methods must include (i) probabilistic aggregation; (ii) behavioral aggregation; (iii) axiomatic aggregation (e.g., linear and log-linear pools); (iv) information theoretic methods (especially for distributed inference); (v) imprecise probabilities and fuzzy logic, as well as upper and lower bounds on probabilities; (vi) graph models such as belief propagation and hybrid inference (involving mixtures of discrete and continuous signals); and (vii) methods of causal reasoning.

The traditional view of information fusion and decision making over sensor networks is heavily biased by the fact that dependencies between information sources are treated only in terms of correlation. Inspired by brains excellent job at processing information, we make a departure from this traditional view point by realizing that: human judgments about the likelihood of events and dependencies among variables are strongly influenced by the perception of cause-effect relationships. However, current man-made systems only employ correlation-type measures of dependencies rather than incorporating causal relationships.

In terms of incorporating causality, we pursued formulation of information-theoretic causality metrics. One such metric is directed information. Unlike mutual information, directed information encompasses dynamics and causality. This project focused on developing a general framework for inferring causal influences in stochastic networks as well as information fusion in online recommendation systems. There were 2 principal thrusts for this project. The first centered around developing the theoretical foundation for identifying causal influences between processes in a network. The second thrust considered the problem of fusion of information in online recommendation systems that use votes from experts (often other users) to recommend objects to customers.

2 Research Results

2.1 Thrust 1: Graphical Models for Representing Causal Influences

In his 1969 paper, Clive Granger, British economist and Nobel laureate, proposed a statistical definition of causality between stochastic processes. It is based on whether causal side information helps in a sequential prediction task. However, his formulation was limited to linear predictors. We proposed a generalized framework, where predictions are beliefs and compare the best predictor with side information to the best predictor without side information. The difference in the prediction performance, i.e., regret of such predictors, is used as a measure of causal influence of the side information. Specifically when log loss is used to quantify each predictors loss and an expectation over the outcomes is used to quantify the regret, we showed that the directed information, an information theoretic quantity, quantifies Granger causality. We also explored a more pessimistic setup perhaps better suited for adversarial settings where minimax criterion is used to quantify the regret. Morevoer, we generalized the notion of cauality to more than a pair of processes. That is we investigated the problem of graphically representing causal influences between processes in a network in a concise manner. To depict causal influences, we developed a probabilistic graphical model analogous to Markov and Bayesian networks, but more meaningful for networks of processes. We showed that this graphical model is equivalent to graphs based on generative models and thus meaningfully summarizes the causal interdependencies, even when there is feedback. We also developed an efficient algorithm for determining the graphical structure of the causal influences when an upperbound for the number of incoming edges for each node is known.

Additionally, we developed an efficient algorithms for finding the best directed approximations for a causal network. Specifically, we considered approximating the true joint distribution on multiple random processes by a tree. That is a graphical model whose directed information graph has at most one parent for any node. Under a Kullback-Leibler (KL) divergence minimization criterion, we showed that the optimal approximate joint distribution can be obtained by maximizing a sum of directed informations. In particular, (a) each directed information calculation only involves statistics amongst a pair of processes and can be efficiently estimated; (b) given all pairwise directed informations, an efficient minimum weight spanning directed tree algorithm can be solved to find the best tree. We demonstrated the efficacy of this approach using simulated and experimental data. In both, the approximations preserve the relevant information for decision- making.

Relatedly, in approximating networks, some networks are not well approximated by a tree. We also investigate the question that whether there are efficient ways to identify best network structure approximations where the approximations is no longer with a tree but has a more complicated structure. The main advantage of such an approximation is the that unlike Inferring the full structure of the generative model which requires calculating divergences using the full joint statistics, finding lower dimensional divergences suffices. For the case when an upperbound on the indegree of each process is known, we have discovered a computationally efficient method using directed information which does not require the full statistics and recovers the parents of each process independently from finding the parents of other processes.

2.2 Thrust 2: Infromation Fusion in Online Recommendation System

Here we studied the problem of fusion of information in online recommendation systems which use votes from experts or other users to recommend objects to customers. We proposed a recommendation algorithm that uses an average weight updating rule and proved its convergence to the best expert and derive an upper bound on its loss. Often times, recommendation algorithms make assumptions that do not hold in practice such as requiring a large number of the good objects, presence of experts with the exact same taste as the user receiving the recommendation, or experts who vote on all or majority of objects. Our algorithm relaxed these assumptions. Besides theoretical performance guarantees, our simulation results showed that the proposed algorithm outperforms current state-of-the-art recommendation algorithm, Dsybil.

We also studied the adversarial setting for information fusion. More precisely, we considered a scenario of learning with expert advice framework in which one of the experts has the intention to compromise the recommendation system by providing wrong recommendations. The problem was formulated as a Markov Decision Process (MDP), and solved by dynamic programming. Somewhat surprisingly, we proved that, in the case of logarithmic loss, the optimal strategy for the malicious expert is the greedy policy of lying at every step. Furthermore, a sufficient condition on the loss function was provided that guarantees the optimality of the greedy policy. Our experimental results however, showed that the condition was not necessary; as for instance, for square loss, the greedy policy was still optimal. Even though, the square loss did not satisfy the condition. Moreover, the experimental results suggested that for absolute loss, the optimal policy is a threshold one.

2.3 Broader Impact

The study of causal influences among processes involved use of point process models and fundamentals from information theory and Bayesian estimation which paved a theoretical path for analysis of information that can be transferred via timing. Examples of such scenarios are timing side channels and active flow linking using timings. We acknowledged the support from this grant in paving for the aforementioned research. One important ramification of this research is that side channels provide nontraditional means of fusing information that must be accounted for in system design. A short summary of the carried out research follows.

Traditionally, scheduling policies have been optimized to perform well on metrics such as throughput, delay and fairness. In the context of shared event schedulers, where a common processor is shared among multiple users, one also has to consider the privacy offered by the scheduling policy. The privacy offered by a scheduling policy measures how much information about the usage pattern of one user of the system can be learnt by another as a consequence of sharing the scheduler. We showed that the most commonly deployed scheduling policy, the first-come-first-served (FCFS) offers very little privacy to its users. Further, we asked the question, is a trade-off between delay and privacy fundamental to the design to scheduling policies? In particular, is there a work-conserving, possibly randomized, scheduling policy that scores high on the privacy metric? Answering the first question, we showed that there does exist a fundamental limit on the privacy performance of a work-conserving scheduling policy. We quantified this limit. Furthermore, answering the second question, we demonstrated that the round-robin scheduling policy (a

deterministic policy) is privacy optimal within the class of work-conserving policies.

Digital fingerprinting is a framework for marking media files, such as images, music, or movies, with user-specific signatures to deter illegal distribution. Multiple users can collude(fuse their information) to produce a forgery that can potentially overcome a finger- printing system. We proposed an equiangular tight frame fingerprint design which is robust to such fusion attacks. We motivate this design by considering digital fingerprinting in terms of compressed sensing. The attack is modeled as linear averaging of multiple marked copies before adding a Gaussian noise vector. The content owner can then determine guilt by exploiting correlation between each users fingerprint and the forged copy. The worst-case error probability of this detection scheme is analyzed and bounded. Simulation results demonstrated the average-case performance is similar to the performance of orthogonal and simplex fingerprint designs, while accommodating several times as many users.

3 Scientific Personnel Supported by This Project

University of Illinois at Urbana-Champaign (Illinois):

1. Negar Kiyavash, Assistant Professor, Illinois-IESE&CSL
2. Anh Truong, Graduate Student, Illinois-ECE&CSL
3. Xun Gong, Graduate Student, Illinois-ECE&CSL

It is notetworthy that Christopher Quinn, a graduate student at ECE department at Illinois also contributed significantly to the project. However, Chris is supported through a Department of Energy Fellowship.

4 Publications Resulting from the Grant

- C. Quinn, N. Kiyavash, and T. Coleman, "Directed Information Graphs," in preparation for submission.
- S. Kadloor and N. Kiyavash, "Delay Optimal Policies Offer Very Little Privacy," submitted to IEEE Transactions on Information Theory, 2013.
- A. Truong and N. Kiyavash, "Learning From Sleeping Experts: Rewarding Informative, Available, and Accurate Experts" submitted, IEEE Transactions on Signal Processing, 2013.
- X. Gong, N. Kiyavash and P. Venkitasubramaniam, "An Information-Theoretic Study of Timing Side Channels in Schedulers," submitted, IEEE Transactions on Information Theory, 2013.
- S. Kadloor and N. Kiyavash, "Delay Optimal Policies Offer Very Little Privacy," to appear in International Conference on Computer Communications (INFOCOM), 2013. (Acceptance rate: 17%)

- C. J. Quinn, N. Kiyavash, and T. P. Coleman, “Efficient methods to compute optimal tree approximations of directed information graphs,” *IEEE Transactions on Signal Processing*, 61(12):3173–3182, June 2013.
- D. Mixon, C. Quinn, N. Kiyavash, and M. Fickus, “Equiangular Tight Frame Fingerprinting codes,” IEEE Transactions on Information Theory, pp. 1855-1865, Vol. 59, No. 3, 2013.
- J. Etesami, N. Kiyavash, and T. Coleman, “Learning Minimal Latent Directed Information Trees,” IEEE International Symposium on Information Theory (ISIT), 2012.
- A. Truong, N. Kiyavash, and V. Borkar, “Convergence Analysis for an Online Recommendation System,” IEEE Conference on Decision and Control (CDC), 2011.
- C. Quinn, N. Kiyavash, and T. Coleman, “Equivalence between Minimal Generative Model Graphs and Directed Information Graphs,” IEEE International Symposium on Information Theory (ISIT), 2011
- D. Mixon, C. Quinn, N. Kiyavash, and M. Fickus, “Equiangular Tight Frame Fingerprinting Codes,” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12-09-2013	Final	7/15/2010-9/12/2013

4. TITLE AND SUBTITLE	
Information Theoretic Causal Coordination	5a. CONTRACT NUMBER FA9550-10-1-0345
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Negar Kiyavash	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Illinois Urbana-Champaign 506 S. Wright Street, 364 Henry Admin Bldg Urbana, IL 61801-3620	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR 875 N. Randolph St. Rm 3112 Arlington, VA 22203	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approval for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the authors and should not be construed as an official U.S. Government position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

Automated decision-making over large-scale distributed systems in the presence of uncertainty and incomplete information is a formidable task. The traditional view of information fusion and decision making over sensor networks is heavily biased by the fact that dependencies between information sources are treated only in terms of correlation. Inspired by brain's excellent job at processing information, we make a departure from this traditional view point by realizing that: human judgments about the likelihood of events and dependencies among variables are strongly influenced by the perception of cause-effect relationships.

In terms of incorporating causality, we pursued formulation of information-theoretic causality metrics such as directed information. We developed a general framework for inferring causal influences in stochastic networks as well as information fusion in online recommendation systems.

15. SUBJECT TERMS

Causal Inference, Information Fusion, Recommendation Systems.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	5	Dr. Negar Kiyavash (217) 244-2525